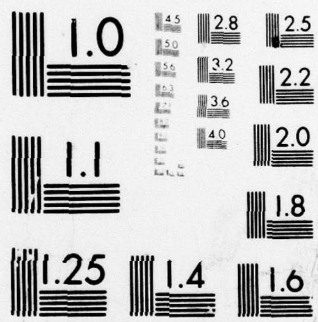
AD-A081 174 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
TIME SERIES ANALYSIS METHODS AND APPLICATIONS: INCREASED INTERD--ETC(U)
NOV 79 E PARZEN N00014-78-C-0599
UNCLASSIFIED TR-N-10 NL

1 OF 1
AD
A081 174

END
DATE FILMED
3-80
DDC




MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

ADA081174

DDC FILE COPY

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

LEVEL

Time Series Analysis Methods and Applications:
Increased Interdisciplinary Interaction
Could Stimulate Research Breakthroughs

by Emanuel Parzen
Texas A&M University

TR-
Technical Report No. N-10
November 1979

11

Texas A & M Research Foundation
Project No. 3838

"Multiple Time Series Modeling and Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

N00014-78-C-0599

DTIC
SELECTED
FEB 28 1980
C
D

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER: Technical Report No. N-10 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle): Time Series Analysis Methods and Applications: Increased Interdisciplinary Interaction Could Stimulate Research Breakthroughs. | | 5. TYPE OF REPORT & PERIOD COVERED: Technical |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s): Emanuel Parzen | | 8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0599 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Code 436, Arlington, VA 22217 | | 12. REPORT DATE: November 1979 |
| | | 13. NUMBER OF PAGES: 20 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report): Unclassified |
| | | 15a. DECLASSIFICATION DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited. | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): NA | | |
| 18. SUPPLEMENTARY NOTES: NA | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Time Series Analysis, Research Directions | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number): This paper consists of three parts. I. (1) Classification of Researchers Applying and Developing Methods of Time Series Analysis. II. 2 Theory of Time Series Analysis. III. 3 Current Fields of Application of Time Series Analysis. | | |

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

80 2 19 132

347380

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS

**Time Series Analysis Methods and Applications:**
**Increased Interdisciplinary Interaction**
**Could Stimulate Research Breakthroughs**

by

Emanuel Parzen
Institute of Statistics
Texas A&M University

## Introduction

The importance (and great potential) for applications of the theory and methods of time series analysis is best evidenced by a catalog of fields to which time series analysis is being applied (See Part III). It is clear that currently extremely diverse research fields are applying time series analysis. The question which I believe is of greatest interest to this panel is whether a significant increase in the quantity, quality, and societal relevance of research in theory and methods of time series analysis could be accomplished through appropriate support by those whose mission is enhancement of the research environment.

To be presented to the National Academy of Sciences Panel on Applied Mathematics Research Alternatives for the Navy in Washington, D.C., November 2, 1979.



My views on the answer to this question are summarized in the following statements:

1. The discipline of time series analysis is enjoying an era of exciting new developments in methods for forecasting, spectral estimation, modelling, parameter estimation and data compression of observed time series.
2. These currently developing methods have the potential to stimulate important insights and research breakthroughs in the study of any scientific or social problems where the variables being studied are evolving in time or space. (A breakthrough is defined to be an ability to solve a problem that could not have been solved otherwise).
3. This potential can be realized by improved interdisciplinary communication between researchers in quite different fields whose data arise as time series or data fields, and increased recognition among the wide scientific community of the potential of the theory and methods of time series analysis for systematic and routine applications.

The outstanding characteristic of time series analysis seems to be that to any problem there is not a single method that should be used for all fields and problem areas; however one can define time series model types, and once one has established from a data set to which model type it belongs there may be "optimal" methods available. There is need to develop the

methods of time series analysis into an integrated framework that provides methods suitable for all the fields of possible application. This would foster interdisciplinary inter-action that would speed the transmission of new insights and methods of time series analysis from one applied field (which originated it) to other applied fields where it could provide scientific breakthroughs.

This paper consists of three parts.

I. Classification of Researchers Applying and Developing Methods of Time Series Analysis.

II. Theory of Time Series Analysis.

III. Current Fields of Application of Time Series Analysis.





## Part I

## Classification of Researchers Applying and Developing Methods of Time Series Analysis

Time series analysis may be defined as a branch of statistical data analysis concerned with "autocorrelated" or statistically dependent data. Such data arises when the data represents a sample path (or realization) of a stochastic process in time or space.

For any branch of statistics, a scientist involved in its theory, methods, and applications may be classified as a

1. statistician, or
2. statistical methods user.

A statistician is a methematical scientist who identifies himself (or herself) as a researcher or consultant on methods for the analysis of data.

A statistical methods user is a scientist, or researcher in a subject matter field, whose professional interest is the quantitative conclusions drawn, rather than the quantitative methods used to infer them.

These two groups can be sub-divided further (See Tables I-A and I-B for definitions).

1. Statisticians
   a. expert
   b. applied
   c. mathematical

d. computational

e. uninformed

2. Statistical Methods Users in a Given Scientific Field

a. expert

b. applied

It would be interesting to take a survey to determine the individuals who belong to the foregoing categories with respect to time series analysis. I conjecture that one would find that the group of applied methods users is very large, and is very eager for improved communication about time series analysis methods.

Among statisticians who are experts in time series analysis one would find further distinctions; experts in the spectral domain, experts in the time domain, experts in both, and experts in all the fields which need to be blended to develop the full range of time series methods:

(1) probability theory to study stochastic models of dynamic processes evolving in time and space,

(2) statistical theory to study methods of inference for estimating parameters and identifying models,

(3) approximation theory to study how to parametrically represent arbitrary functions, especially positive functions,

(4) numerical analysis to study methods of handling large masses of data and efficient computer algorithms, and

(5) systems theory to study the representation of time series as inputs and outputs of systems.



The theory of time series analysis lies at the intersection of the mathematical, statistical, computational, and system sciences, and provide an elegant interplay among these disciplines. It provides the means of applying advanced mathematical ideas and theorems to contribute towards the solutions of very practical problems. Time series analysis is truly an interdisciplinary field, because development of its theory and methods requires interaction between the diverse disciplines in which it is being applied. To harness its great potential there must develop a community of statistical and other scientists who are educated and motivated to have a background in the theory and methods of time series analysis adequate to handle the problems of time series analysis in all the fields in which they occur.

I believe that currently in the time series literature there are many flowers that "bloom unseen". What is needed is an interested community large enough so that when important articles are published there would be many "to applaud, to contest, to amplify, and to develop it in new dimensions and new directions, and to find new applications" (The quotation is from an article in the Sunday New York Times, October 21, 1979, p. E9, by Gardner Ackley entitled "Why is American Economics Different and Distinguished?").

A Proposal for a Time Series Conference Board

One way to stimulate a time series community might be to form an interdisciplinary committee which would organize an annual workshop, perhaps involving different types of audiences each year. A possible format for these workshops is as follows: (1) issue a call for papers on the theme of the workshop (applications and/or methods), (2) select papers which would receive an award of $1500, (3) the papers would be presented and intensively discussed in a half day session at the workshop, (4) a proceedings of the workshop would publish the papers and discussion.

A Steering Committee for the workshops could be selected by inviting outstanding researchers, and also by asking relevant scientific societies to select a representative. The name of the steering committee could be the "Conference Board of Time Series Methods and Application".

In addition to organizing an annual workshop, as described above, the Conference Board could have members who would receive a regular newsletter with news of conferences and publications in time series methods and applications, and also short announcements of research results. An annual budget of about $50,000 would be required, mainly to fund travel to the workshop and to Conference Board meetings.



Table I-A

Levels of Involvement of Statisticians in a Branch of Statistics (such as Time Series Analysis)

| | |
|---|---|
| Expert: | Statistician who has a broad education in the theory of the branch of statistics, whose research is oriented to the development of new methods, and who has an excellent command of the recognized methods. |
| Applied: | Statistician who has a good command of recognized methods of the branch of statistics, and who has the ability to collaborate and consult with researchers in subject matter fields who would like to use time series analysis in their research. |
| Mathematical: | Statistician whose research is oriented to the mathematical properties of methods used or proposed for time series analysis. |
| Computational: | Statistician whose specialty is statistical computing and who implements in computer programs new methods of time series analysis. |
| Un-informed: | Statistician who is not educated enough in the methods of the branch of statistics to be able to offer any counsel about how to apply them. |

Table I-B

Levels of Achievement of Statistical Methods Users
Whose Primary Identification is as Researchers
in a Scientific Field in Which Time Series Problems Arise

<u>Expert Methods User</u>. Scientist who has an excellent command of theory and methods of a branch of statistics, and is the first to apply (and possibly originate) new methods in his subject matter area.

<u>Applied Methods User</u> Scientist who has a good command of, and uses in his research, recognized methods of a branch of statistics which have been applied in his (or her) subject matter field.



## Part II

## Theory of Time Series Analysis

Some major theoretical problems of time series analysis which have been considered by statisticians are listed in Tables A, B, and C. The basic problems in Tables A and B are called the stage of "linear Gaussian analysis" because the models assumed involve linear relationships between normally distributed random variables, and use mean square error estimation criteria.

Through implementation in computer programs, methods of time series analysis can be transmitted to statistical methods users who are unable to follow theoretical justifications. Illustrations of the available computer software are provided by TIMSAC-78 (a library of subroutines developed by Professor Hirotugu Akaike of the Institute of Statistical Mathematics, Tokyo, Japan) and TIMESBOARD (a time series subroutine library implementing time series research directed by E. Parzen and developed by H.J. Newton). Other time series researchers also have prepared extensive computer subroutines (for example G.E.P. Box, G.M. Jenkins, G.C. Tiao, R.H. Jones, H.L. Gray, D.Findley, and B.W. Dickinson). It would be of value to maintain a central file of information about time series algorithms which have been implemented by time series researchers.

From a practical point of view, the problems of linear Gaussian analysis listed in Tables I and II are not definitively solved. In my opinion, <u>a high priority need is for research</u> which evaluates and compares the available methods for estimation

and testing of linear Gaussian models with the aim of arriving at definitive data adaptive methods of empirical time series analysis.

It is clear that no estimation or model identification procedure can be successfully applied automatically to every set of data. But I believe one can develop methods of classification of observed time series into model types which can distinguish the cases where routine statistical analysis is applicable from cases requiring more "custom tailoring". (At Texas A&M we are currently engaged in illustrating such an approach in a paper entitled "Forecasting, and Time Series Model Types of, 101 Economic Time Series".)

As illustrations of the kinds of open questions of linear Gaussian analysis which are very important to applied researchers, and which deserve investigation as a part of an overall integration and consolidation of theory and methods in time series analysis, I would like to mention the following (which are currently being investigated):

1. Is it useful to regard the ARIMA models introduced by Box and Jenkins as a generalization of ARMA models? (ARMA models are appropriate for short memory time series, while the I part should be regarded as an attempt to find a transformation of a long memory time series to a short memory one);



2. Box-Jenkins seasonal models usually involve long order moving average components; one should investigate by empirically analyzing many observed time series whether such models would be identified as the best fitting models using other approaches (such as the S-Arrays of H.L. Gray and colleagues).
3. Even the question of what estimation methods to use for the parameters of an ARMA(p,q) scheme is not clearly resolved in my opinion, especially if the parameters are not of interest for their own sake, but are only coefficients in prediction formulas. So-called maximum likelihood estimators may not then be the preferred estimators.
4. What are the relations between autoregressive and maximum entropy spectral estimators? What is the effect of the choice of autoregressive order (or prediction error filter length)?
5. Can criteria be developed to indicate how predictable a time series is?
6. How to formulate state space models so that they are easily transformed to equivalent ARMA models?

## Table A

### Linear Gaussian Analysis of a Single Time Series

1. Testing hypothesis of white noise, (that is, data $Y(t)$ has no memory).

   Distribution theory of test statistics, especially those derived against the alternative hypothesis of hidden periodicities.
2. Efficient estimation of parameters of an ARMA(p,q) model,

   Distribution theory of parameter estimators.
3. Identification methods for determining the orders p and q.
4. Estimation of the spectral density, assuming data $Y(t)$ is stationary time series with short memory (defined intuitively here as summable correlation function $\rho(v)$),

   Distribution theory of spectral estimators.
5. Estimation of parameters of model $Y(t) = S(t) + Z(t)$, where $S(t)$ is a sum of harmonics of unknown frequencies and $Z(t)$ is short memory stationary time series.

   Mixed spectral analysis.
6. Estimation of parameters of model $Y(t) = \mu(t) + Z(t)$ where $\mu(t)$ obeys a regression model and $Z(t)$ is short memory

   Regression with correlated error.
7. Decomposition of a time series into a sum of seasonal, trend, and stationary error components.
8. Forecasting or prediction of stationary time series.
9. Modelling and forecasting of non-stationary time series.
10. State space modelling and forecasting of time series.



## Table B

### Linear Gaussian Analysis of Multiple Time Series

1. Testing hypothesis of incoherence (all cross-correlations are zero).
2. Efficient estimation of parameters of multiple ARMA(p,q) model.
3. Identification methods for determining orders p and q.
4. Estimation of spectral density matrix.
5. Estimation of coherence, phase, and transfer functions in the frequency domain.
6. Estimation of transfer functions (input-output representation) in the time domain.
7. Canonical analysis of multiple time series; dimensionality reduction.
8. Forecasting or prediction of multiple stationary time series.

## Table C

**Time Series Problems of Current Research Interest to Statisticians other than Data Adaptive Formulations of Linear Gaussian Methods.**

1. Missing or irregularly observed data.
2. Bad data and robust inference.
3. Non-Gaussian Models (especially infinite variance).
4. Transformation of the data.
5. Non-linear models.
6. Time-varying parameters.
7. Point processes (series of events).
8. Spatial processes: modelling.
9. Spatial processes: interpolation
10. Continuous time processes.
11. Reproducing Kernel Hilbert space models.
12. Signal Detection.



# Part III

## Current Fields and Journals of Application of Time Series Analysis

In order to understand the current state of development, and the future potential, of applications of time series analysis we attempt to classify the various fields of application of, and list some journals in which are published papers on methods and applications of time series analysis.

1. Physical and Earth Sciences

   Geophysics, seismology, oceanography, geology, astronomy, atmospheric physics, meteorology.

2. Electrical Engineering Signal Processing

   Acoustics, speech, sonar, radar, communications, image reconstruction, pattern recognition.

3. Engineering other

   Aerospace guidance, structural vibration studies, system identification, hydrology, process identification.

4. Biological Sciences

   Biorhythms.

5. Medicine

   EEG and EKG analysis, monitoring of physiological variables.

6. Economics and Management Sciences

   Business cycles, econometrics, forecasting of micro-economic and macroeconomic variables, finance.

7. Social Sciences

   History, political science, psychology, sociology.

8. Statistical and Computational Sciences

   The methods of time series analysis provide ways of thinking for solving problems which do not arise as

the modeling of fields of data in time or space; they are being applied to non-parametric methods of statistical data analysis, and to numerical solutions of integral equations.

Some Journals in which Time Series Papers Appear.

1. <u>Physical and Earth Sciences</u>

   Rev. Geophysics
   Geophysics
   Journal Geo. Res.
   Geophy. JR Astr. Soc.
   Physics Earth and Planetary Interiors
   Reviews of Geophysics and Space Physics
   Astron. Astrophysics
   J. Phys. Oceanography
   Tellus
   J. Atmospheric Sci.
   Monthly Weather Rev.
   J. Applied Meteorology

2. <u>Electrical Engineering Signal Processing</u>

   IEEE Trans. Information Theory
   IEEE Trans. Circuits and Systems
   IEEE Trans. Acoustics, Speech, Signal Processing
   IEEE Trans. Geoscience Electronics
   IEEE Trans. Automatic Control
   IEEE Trans. Communcations
   IEEE Proceedings



   IEEE International Convention Acoustics, Speech, Signal Processing
   IEEE Conference Decision and Control
   Bell System Technical Journal

3. <u>Engineering other</u>

   Transactions, American Society Mechanical Engineering
   Water Resources Research
   Journal of Hydrology
   J. Hydraulic Amer. Soc. Civil Engineers
   Applied Mathematical Modelling

4. <u>Biological Rhythms</u>

   J. Interdisciplinary Cycle Research
   J. Chronobiology
   Nature

5. <u>Medicine</u>

   IEEE Transactions Bio-medical Engineering
   EEG and Clin. Neurophysiology

6. <u>Economics</u>

   J. Econometrics
   Econometrica
   International Economic Review

7. <u>Social Science, Law and Education</u>

   Law and Society Review
   Psychological Bulletin
   Applied Behavior Analysis
   Evaluation Quarterly
   Sociological Methods and Research
   J. Cycle Research

8. Statistical and Computational Sciences

Biometrika

Jour. Royal Statistical Society

Applied Statistics

Journal of the American Statistical Assoc.

Annals of Statistics

Communications in Statistics

Scand. J. Statistics

Australian J. Statistics

Theory of Probability and its Applications

(Translation of Russian Journal)

J. Applied Probability

Ann. Inst. Stat. Math.

International Statistical Review

Journal of Association for Computing Machinery

Metrika

Technometrics

9. Workshop Proceedings

RADC Spectral Estimation Workshops, 1978, 1979.

IMS Time Series Special Topics Meeting

NATO Signal Processing Advanced Study Seminar

Multivariate Analysis (ed. Krishnaiah)

Tulsa Applied Time Series (ed. Findley)

Census Workshop on Seasonal Analysis of Economic Time Series

10. Books

Under preparation is a list of all English language statistical books on time series analysis, arranged chronologically, and reproducing from each book its title page, preface, and table of contents.



Appendix: Some Recent Publications on Time Series Analysis
by Emanuel Parzen


1977 "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes" Multivariate Analysis - IV, edited by P. Krishnaiah, North Holland: Amsterdam, 283-295.

1979 "An Approach to Modeling Seasonally Stationary Time Series" (with M. Pagano), Journal of Econometrics, 9, 137-153.

1979 "Forecasting and Whitening Filter Estimation" TIMS Studies in the Management Sciences, 12 (1979) 149-165.

1980 "Multiple Time series modeling, II" (with H. J. Newton) Multivariate Analysis - V, edited by P.Krishnaiah, North Holland: Amsterdam.

1980 "Time Series Modeling, Spectral Analysis, and Forecasting" Directions in Time Series Analysis, ed. D. R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics.

1980 "Time Series Model Identification and Prediction Variance Horizon" Second Tulsa Symposium on Applied Time Series Analysis, March.

1980 "Modern Spectrum Analysis", NATO Advanced Study Institute Signal Processing, August.